

Validity, reliability and minimal detectable change of the Balance Evaluation Systems Test (BESTest), Mini-BESTest and Brief-BESTest in patients with end-stage renal disease

Cristina Jácome^{1,2}; Inês Flores²; Filipa Martins²; Conceição Castro³; Charlotte C McPhee⁴; Ellen Shepherd⁴; Sara Demain⁴; Daniela Figueiredo^{2,5}; Alda Marques^{1,2,6}

1 Lab 3R – Respiratory Research and Rehabilitation Laboratory, School of Health Sciences, University of Aveiro (ESSUA), Aveiro, Portugal

2 School of Health Sciences, University of Aveiro (ESSUA), Aveiro, Portugal

3 CentroDial, Centro de Hemodiálise S.A., São João da Madeira, Portugal

4 Faculty of Health Sciences, University of Southampton, United Kingdom

5 Center for Health Technology and Services Research (Cintesis.UA), University of Aveiro, Aveiro, Portugal

6 Institute for Research in Biomedicine (iBiMED), University of Aveiro, Aveiro, Portugal

Corresponding author: Alda Marques, PT, MSc, PhD, Senior Lecturer, Lab 3R – Respiratory Research and Rehabilitation Laboratory, School of Health Sciences, University of Aveiro (ESSUA), Agras do Crasto - Campus Universitário de Santiago, Edifício 30, 3810-193 Aveiro, Portugal. Telephone number: 00351 234 372 462. Fax number: 00351 234 401 597. Email: amarques@ua.pt

Word count: 2625

Validity, reliability and minimal detectable change of the Balance Evaluation Systems Test (BESTest), Mini-BESTest and Brief-BESTest in patients with end-stage renal disease

Purpose: This study determined the validity, test-retest reliability and minimal detectable change of the Balance Evaluation Systems Test (BESTest), Mini-BESTest and Brief-BESTest in patients with end-stage renal disease.

Methods: A cross-sectional study with 74 patients with end-stage renal disease (male 66.2%; 63.9±15.1 years old) was conducted. Participants were asked to report the number of falls during the previous 12 months and to complete the Activity-specific Balance Confidence Scale. The BESTest was administered, and the Mini-BESTest and Brief-BESTest scores were computed based on the BESTest performance. Validity was assessed by correlating balance tests with each other and with the Activity-specific Balance Confidence Scale. Test-retest relative reliability and agreement were explored with the Intraclass correlation coefficient equation (2,1) and the Bland and Altman method. Minimal detectable changes at the 95% confidence level were established.

Results: Balance test scores were significantly correlated with each other (spearman's correlation=0.89-0.92) and with the Activity-specific Balance Confidence Scale (spearman's correlation=0.49-0.59). Balance tests presented high test-retest reliability (Intraclass correlation coefficient=0.84-0.94), with no evidence of bias. Minimal detectable change values were 10.8 (expressed as a percentage 13.5%), 5.3 (23.7%) and 5.6 (34%) points for the BESTest, Mini-BESTest and Brief-BESTest, respectively.

Conclusions: All tests are valid and reliable to assess balance in patients with end-stage renal disease. Nevertheless, based on the minimal detectable changes found, BESTest and Mini-BESTest may be the most recommended tests for this specific population.

1

2 **Key-words**

3 Chronic Kidney Failure; Dialysis; Postural Balance; risk assessment; Reproducibility of

4 Results.

Introduction

Chronic kidney disease affects 11-13% of the population worldwide,[1] being an important public health issue with high socioeconomic burden. This long term and progressive condition is associated with fatigue, muscle weakness and reduced physical activity.[2, 3] As a result, patients with chronic kidney disease, and particularly with end stage renal disease (ESRD) may experience difficulties in performing activities of daily living that require balance control and be at high risk of falling.[4]

Recent literature indicates that 17.2% to 47% of patients with ESRD fall at least once during a 3-12 month period.[5, 6] Thus, valid, reliable and clinically feasible tests for comprehensive assessment of balance impairments are needed to screen balance impairments, inform the design and evaluate the impact of optimal interventions for fall reduction.

A number of balance tests have been described in the literature.[7, 8] The Balance Evaluation Systems Test (BESTest) was developed to assess balance comprehensively. It assesses functioning of six balance control systems[9] and has been one of the most commonly used test in patients with chronic diseases, such as cancer,[10] Parkinson's disease[11] and chronic obstructive pulmonary disease.[12] However, while the clinimetric properties of this test, such as validity, test-retest reliability and minimal detectable change, have been established in several specific populations,[13, 14, 15] they have not yet been investigated in patients with ESRD.

Due to the time required to complete the BESTest (approximately 20-30 minutes),[14] shortened versions were developed, the Mini-BESTest[16] and the Brief-BESTest.[14] The Mini-BESTest includes important aspects of dynamic balance control, reflecting balance challenges during activities of daily living (administration takes approximately

1 15 minutes).[16] In contrast, the Brief BESTest contains items that assess all balance
2 systems outlined by the original BESTest, requires less administration time
3 (approximately 10 minutes) and less equipment, which can favor its clinical use.[14]
4 These shortened versions have also gained interest to assess balance in patients with
5 Parkinson's disease,[17] multiple sclerosis,[14] and balance disorders.[18] Nevertheless,
6 to the authors' knowledge, neither the Mini-BESTest nor the Brief-BESTest have been
7 applied, or their clinimetric properties studied, in patients with ESRD. Determining the
8 clinimetric properties and associated measurement error of these tests is fundamental to
9 decide whether they are appropriate to detect balance impairments in patients with
10 ESRD[19] and can be used in the clinical management of these patients.
11 Therefore, the purpose of this study was to determine the validity, test-retest reliability
12 and the minimal detectable change of the BESTest, Mini-BESTest and the Brief-BESTest
13 in patients with ESRD.

14 **Methods**

15 *Study design and ethics*

16 A cross-sectional study was conducted from December 2013 to June 2014 in the central
17 region of Portugal. Ethical approval was obtained from the Ethics Committee of the
18 Research Unit of Health Sciences at the School of Nursing in Coimbra (P185-10/2013).
19 This study is reported following the COnsensus-based Standards for the selection of
20 health Measurement INstruments,[20] the STrengthening the Reporting of OBservational
21 studies in Epidemiology statement[21] and the guidelines for reliability and agreement
22 studies.[19]

23 *Participants*

Three renal dialysis centers participated and clinicians in each center identified eligible patients according to the following criteria: (i) had ESRD; ii) were receiving adequate hemodialysis or peritoneal dialysis for at least 3 months ($Kt/V > 1.2$) and iii) were aged 18 years or older. Patients were excluded if they i) had an hospitalization within the past 3 months due to worsening of their health status; ii) had comorbidities that interfered with independent ambulation (e.g., hip fracture, lower limb amputation); iii) were taking medication that could cause dizziness or affect their balance (e.g., psychotropic medications); iv) had severe respiratory (e.g., chronic obstructive pulmonary disease), neurological (e.g., Parkinson's disease, multiple sclerosis), musculoskeletal (e.g., severe osteoarthritis) or psychiatric (e.g., psychosis, schizophrenia) disorders, that could interfere with the measurements. After screening, patients were contacted by the clinician, who explained the purpose of the study and asked about their willingness to participate. When patients agreed to participate, an appointment with the researchers was scheduled at their dialysis center. Written informed consent was obtained prior to data collection.

Data collection procedures

Two research physiotherapists collected the data. Socio-demographic (age and gender), anthropometric (height, weight, body mass index) and clinical (self-reported comorbidities - e.g., hypertension and hyperlipidemia) data were first collected. Patients were then provided with a clear definition of falls (an event when you find yourself unintentionally on the ground, floor or lower level)[22] and required to report their history of falls using two standardized questions (1. "*Have you had any falls in the last 12 months?*" and, if yes, 2. "*How many times did you fall down in the last 12 months?*").[23] The self-reported Activities-Specific Balance Confidence scale was used to assess balance confidence.[24] This scale quantifies an individual's perceived ability to maintain

his/her balance under different circumstances, using a scale of 0% (no confidence) to 100% (total confidence).[25] Finally, the BESTest was performed and participants were encouraged to rest, as needed.

For each item of the BESTest, the physiotherapist read the instructions to the participant and performed the task. Then, the participant performed the task with close supervision.

For test-retest reliability, participants were reassessed by 1 of the 2 physiotherapists within 5 days after session 1. Effort was made to keep all factors associated with the testing sessions consistent, specifically the time of day, area in which the tests were performed and use of a walking aid (if needed). The Mini-BESTest and Brief-BESTest scores were computed based on the performance of the BESTest tasks. A custom designed worksheet was used by the physiotherapists to simultaneously record the BESTest and the Mini-BESTest item scores. Brief-BESTest scores were extracted from the relevant subset of BESTest items.

Balance Tests

BESTest

The BESTest contains 36 items organized into 6 subsections: biomechanical constraints, stability limits and verticality, anticipatory postural adjustments, postural responses to external perturbations, sensory orientation during stance and stability in gait.[9] Each item is scored from 0 (severe balance impairment) to 3 (no balance impairment) and the maximum possible total score is 108 points. BESTest has high test-retest reliability (Intraclass Correlation Coefficient - ICC=0.87-0.92) in individuals with and without balance disorders, in patients with Parkinson's disease, in older cancer survivors and in patients with chronic obstructive pulmonary disease.[9, 10, 13, 15]

Mini-BESTest

The Mini-BESTest includes 14 items from sections of the BESTest related to anticipatory postural adjustments, reactive postural responses, sensory orientation and stability in gait.[16] Two of the 14 items are scored bilaterally. Each item is scored from 0 (severe balance impairment) to 2 (no balance impairment) and the maximum possible score is 28 points. Higher scores indicate better balance performance. High test-retest reliability (ICC=0.88-0.97) have also been found in patients with balance disorders, Parkinson's disease, chronic stroke, chronic obstructive pulmonary disease and also in older cancer survivors.[10, 11, 15, 18, 26]

Brief-BESTest

The Brief-BESTest is a 6-item balance test, which contains 1 item from each of the 6 subsections of the BESTest.[14] Two of the 6 items are scored bilaterally, resulting in an 8-item balance test. Each item is scored from 0 (severe balance impairment) to 3 (no balance impairment) and the maximum possible score is 24 points. Higher scores indicate better balance performance.[14] It has also showed high test-retest (ICC=0.82-0.94) reliability in patients with chronic obstructive pulmonary disease,[15] older cancer survivors[10] and older adults living in the community.[27]

Statistical analysis

Descriptive statistics were used to describe the sample. A z-test was applied for normality test using skewness and kurtosis.[28] Patients with a history of falls were defined as those who reported at least one fall during the past year; patients without a history of falls were defined as those who reported no falls during the past year. Participants' characteristics were compared between i) those with and without a history of falls and ii) those included in the reliability analysis and the remaining sample, using independent t-tests for normally

distributed data, Mann-Whitney U-tests for non-normally distributed and Chi-square tests for categorical data.

Reliability

Test-retest reliability was analyzed in a subsample of the first consecutive 31 participants. This sample size was determined according to the study from Bonnet,[29] which has established that a minimum of 21 individuals were necessary to estimate an ICC of 0.9 with a 95% confidence interval width of 0.230 ($\alpha=0.05$ and $k=2$). As interventions with patients with chronic kidney disease at various stages have considerable dropouts,[30] a 40% attrition rate was estimated, yielding a sample of 31 participants.

Test-retest reliability was computed using the scores from the same physiotherapist in sessions 1 and 2. For relative reliability, the ICC equation (2,1) was used as it was intended to generalize the present results to a variety of raters[31]. ICC was interpreted as excellent (>0.75), moderate to good ($0.4-0.75$) or poor (<0.4).[32] Agreement was explored with the Bland and Altman method and standard error of measurement (SEM). The Bland and Altman method was used for visual judgement of how well the measurements between sessions agreed.[33]

The SEM indicates the extent to which a score varies on repeated measurements. It provides a value for measurement error in the same units as the measurement itself and thus, is easier to use in clinical practice.[34] It was calculated using equation 1:

$$SEM = SD \sqrt{(1 - ICC)} \quad (1)$$

where SD is the standard deviation of the scores obtained from all individuals and ICC is the test-retest reliability coefficient.

Minimal Detectable Change

The MDC at the 95% level of confidence (MDC_{95}), is also reported in the same units as the measurement itself and represents the smallest change that can be interpreted as a real difference. It was calculated as follows (equation 2):

$$MDC_{95} = SEM \times 1.96 \times \sqrt{2} \quad (2)$$

The MDC was also expressed as a percentage ($MDC_{\%}$), defined as (equation 3):

$$MDC_{\%} = (MDC_{95}/mean) \times 100 \quad (3)$$

where mean is the mean of the scores obtained in the 2 testing sessions. The $MDC_{\%}$ is independent of the units of measurement and facilitates the comparison of the random measurement error among different measures.[35] A $MDC_{\%}$ below 30% is considered acceptable.[36]

Validity

Spearman's correlation coefficient (ρ) was used to examine the relationship among balance tests (concurrent validity) and between each balance test and the Activities-specific Balance Confidence scale (convergent validity).

All statistical analyses were performed using IBM SPSS Statistics version 20.0 (IBM Corporation, Armonk, NY, USA) and plots created using GraphPad Prism version 5.01 (GraphPad Software, Inc., La Jolla, CA, USA). The level of significance was set at 0.05.

Results

Participants

A total of 131 patients were screened and 82 were eligible. Two participants refused to participate and 6 did not complete the assessment. Therefore, 74 participants were included. Participants had a mean age of 63.9 ± 15.1 years old and were mainly male ($n=49$; 66.2%). Twenty-eight (37.8%) participants had history of falls. Twelve

participants used walking aids, eleven used a cane and one used a walker. Participants' characteristics and balance scores are presented in table 1. All balance tests were able to significantly differentiate between participants among age ranges (Table 2).

(table 1)

(table 2)

Validity

All balance tests were strongly correlated with each other, with ρ ranging from 0.89 to 0.92 ($p < 0.001$). The Activities-specific Balance Confidence scale was significantly correlated with the BESTest ($\rho = 0.59$), Mini-BESTest ($\rho = 0.49$) and the Brief-BESTest ($\rho = 0.5$) (figure 1).

(figure 1)

Test-retest reliability

The characteristics of participants included in the reliability analysis ($n = 31$) were not significantly different from remaining participants. Table 2 presents the relative test-retest reliability and the agreement of the BESTest, Mini-BESTest and the Brief-BESTest. Excellent test-retest reliability was observed. Good test-retest agreement was verified for all balance tests, with mean differences close to zero (table 3 and figure 2).

(table 3 and figure 2)

Minimal detectable change

The MDC_{95} was 10.8 ($SEM = 3.9$; $MDC\% = 13.5\%$), 5.3 ($SEM = 1.9$; $MDC\% = 23.7\%$) and 5.6 ($SEM = 2$; $MDC\% = 34\%$) for the BESTest, Mini-BESTest and the Brief-BESTest, respectively.

Discussion

1 To our knowledge, this is the first study to investigate the validity, test-retest reliability
2 and minimal detectable change of the BESTest, Mini-BESTest and the Brief-BESTest in
3 patients with ESRD. The three tests are valid and reliable to assess balance in patients
4 with ESRD. The preliminary MDC found for these balance tests can be used by clinicians
5 to identify a true change in balance in patients with ESRD.

6 In this study, 37.8% of patients reported at least one fall in the previous 12 months. This
7 result is in line with previous research.[5, 6] In healthy older adults slightly lower
8 prevalences have been described (12.1%-33%).[27, 37, 38] In addition, patients with
9 history of falls were older than those without history of falls and had worse performance
10 in balance tests. It is known that older adults frequently present reduced skeletal muscle
11 strength, gait speed and physical activity levels [39]. These impairments may have also
12 contributed to the balance deficits found in patients with ESRD. This finding highlights
13 the need of screening for balance impairments in the comprehensive assessment of
14 patients with ESRD, particularly in those with more advanced age.

15 The BESTest, Mini-BESTest, and Brief-BESTest have shown to be valid in a variety of
16 clinical populations.[10, 18, 26, 40] In patients with ESRD, the three balance tests were
17 also significantly associated with each other and with the Activities-specific Balance
18 Confidence scale, demonstrating good concurrent and convergent validity.

19 Excellent results were found for test-retest reliability of the balance tests (ICC=0.84-
20 0.94). Similar results have been found in reliability studies conducted in other
21 populations, such as Parkinson's disease (ICC=0.88-0.90),[13] older cancer survivors
22 (ICC=0.90-0.94),[10] chronic obstructive pulmonary disease (ICC=0.82-0.87)[15] and
23 stroke (ICC=0.97).[26] Regarding agreement, results were also encouraging since
24 balance tests showed good agreement, with no systematic bias. Thus, it seems that

1 clinicians can be confident in using these three balance tests to assess balance
2 impairments in patients with ESRD.

3 In this study, preliminary MDC₉₅ values for BESTest, Mini-BESTest and Brief-BESTest
4 were determined, which will be useful for clinicians working with patients with ESRD in
5 assessing whether an intervention (e.g., rehabilitation program) has caused any real
6 change in balance. However, MDCs were slightly higher than those established in other
7 populations: BESTest (10.8 vs. range 6.2-9[10, 15, 41]), Mini-BESTest (5.3 vs. range
8 2.4-3.8[10, 15, 18, 41]) and Brief-BESTest (5.6 vs. range 2.6-4.9[10, 15, 41]). These
9 differences may be explained by the specificities of each population. But may also be
10 related with the samples used. In the present study, participants' mean age was 63.9 years
11 old, with large age range 30-95. In the reported studies patients were somewhat older
12 (mean ages between 66 and 76 years old [10, 15, 18, 27, 41]). It should be also noted that
13 the MDC% of the Brief-BESTest was slightly higher (34%) than the value considered
14 acceptable (i.e., <30%). However, this MDC% was obtained from one single performance
15 at each session. Future studies should determine the MDC₉₅ and MDC% from the average
16 of two performances at each session. This procedure may reduce the magnitude of
17 measurement error and consequently achieve lower values for the MDC₉₅ and MDC%.
18 This has been observed in tests for functional balance assessment, such as the Timed Up
19 and Go test.[42]

20 There are limitations in the present study that need to be acknowledged. First, the small
21 sample size. Yet, the number of patients included is within the range (28 to 122) used in
22 previous research assessing the clinimetric properties of the BESTest, Mini-BESTest or
23 of the Brief-BESTest.[10, 18, 27, 41] Second, the sample was mainly composed of
24 patients aged 60 or over (>70%), so generalization of findings are hindered to the younger

1 population of patients with ESRD. Future studies may replicate the study in larger
2 samples of younger patients and report results per age ranges. Third, patients were
3 classified as having or not having a history of falls based on their self-report hence, some
4 degree of bias in their responses might have been present. Finally, the Mini-BESTest (14
5 tasks) and Brief-BESTest (6 tasks) scores were derived from the patients' performance
6 on BESTest (36 tasks). The length of the BESTest may have produced a certain degree
7 of fatigue. Thus, it might be thought that the scores of the BESTest shortened versions
8 could be slightly different if performed independently from the BESTest. However, in
9 order to address this concern, participants were given frequent resting breaks during the
10 balance assessment. To clarify this, future studies should assess the validity, test-retest
11 reliability and minimal detectable change of the Mini-BESTest and of the Brief-BESTest
12 when performed separately from the BESTest.

13 The BESTest, Mini-BESTest, and the Brief-BESTest are valid and reliable to
14 comprehensively assess balance in patients with ESRD. Nevertheless, based on the
15 minimal detectable changes, Balance Evaluation Systems Test and Mini-Balance
16 Evaluation Systems Test may be the most recommended; and the selection of one of them
17 may be based on time and equipment availability.

18 **Funding**

19 This research received no specific grant from any funding agency in the public,
20 commercial, or not-for-profit sectors.

22 **Conflict of Interest Statement**

23 The Author(s) declare(s) that there is no conflict of interest

References

1. Hill NR, Fatoba ST, Oke JL, et al. Global Prevalence of Chronic Kidney Disease – A Systematic Review and Meta-Analysis. PLoS ONE. 2016;11(7):e0158765. doi: 10.1371/journal.pone.0158765.
2. Workeneh BT, Mitch WE. Review of muscle wasting associated with chronic kidney disease. Am J Clin Nutr. 2010 Apr;91(4):1128S-1132S. doi: 10.3945/ajcn.2010.28608B. PubMed PMID: 20181807; eng.
3. Jhamb M, Liang K, Yabes J, et al. Prevalence and Correlates of Fatigue in Chronic Kidney Disease and End-Stage Renal Disease: Are Sleep Disorders a Key to Understanding Fatigue? Am J Nephrol. 2013;38(6):489-495.
4. Cruz J, Marques A, Jácome C, et al. Global Functioning of COPD Patients With and Without Functional Balance Impairment: An Exploratory Analysis Based on the ICF Framework. COPD. 2015 Aug 5;12(2):207-16. doi: 10.3109/15412555.2014.933793. PubMed PMID: 25093384; Eng.
5. Galvão MH, Santos LIS, Barcelar JdM, et al. Assessment of the capacity of the falls efficacy scale international in specifying the risk of falling in patients with chronic kidney disease on hemodialysis. Fisioterapia e Pesquisa. 2013;20(2):151-157.
6. Cook WL, Tomlinson G, Donaldson M, et al. Falls and Fall-Related Injuries in Older Dialysis Patients. Clin J Am Soc Nephrol. 2006 November 1, 2006;1(6):1197-1204. doi: 10.2215/cjn.01650506.
7. Sibley KM, Beauchamp MK, Van Ooteghem K, et al. Using the Systems Framework for Postural Control to Analyze the Components of Balance Evaluated

in Standardized Balance Measures: A Scoping Review. Arch Phys Med Rehabil.
2015 1//;96(1):122-132.e29. doi: <http://doi.org/10.1016/j.apmr.2014.06.021>.

8. Oliveira CC, Lee A, Granger CL, et al. Postural control and fear of falling
assessment in people with chronic obstructive pulmonary disease: a systematic
review of instruments, international classification of functioning, disability and
health linkage, and measurement properties. Arch Phys Med Rehabil. 2013
Sep;94(9):1784-1799 e7. doi: 10.1016/j.apmr.2013.04.012. PubMed PMID:
23632285; eng.

9. Horak FB, Wrisley DM, Frank J. The Balance Evaluation Systems Test (BESTest)
to differentiate balance deficits. Phys Ther. 2009 03/10/received
01/30/accepted;89(5):484-498. doi: 10.2522/ptj.20080071. PubMed PMID:
PMC2676433.

10. Huang MH, Miller K, Smith K, et al. Reliability, validity, and minimal detectable
change of Balance Evaluation Systems Test and its short versions in older cancer
survivors: a pilot study. J Geriatr Phys Ther. 2015 Feb 18;39(2):58-63. doi:
10.1519/jpt.0000000000000047. PubMed PMID: 25695466; Eng.

11. Leddy AL, Crowner BE, Earhart GM. Utility of the Mini-BESTest, BESTest, and
BESTest sections for balance assessments in individuals with Parkinson disease.
J Neurol Phys Ther. 2011 Jun;35(2):90-7. doi: 10.1097/NPT.0b013e31821a620c.
PubMed PMID: 21934364; PubMed Central PMCID: PMC3178037. eng.

12. Beauchamp MK, Janaudis-Ferreira T, Parreira V, et al. A randomized controlled
trial of balance training during pulmonary rehabilitation for individuals with
COPD. Chest. 2013 Dec;144(6):1803-10. doi: 10.1378/chest.13-1093. PubMed
PMID: 23975185; eng.

- 1 13. Leddy AL, Crowner BE, Earhart GM. Functional gait assessment and balance
2 evaluation system test: reliability, validity, sensitivity, and specificity for
3 identifying individuals with Parkinson disease who fall. *Phys Ther.* 2011
4 Jan;91(1):102-13. doi: 10.2522/ptj.20100113. PubMed PMID: 21071506;
5 PubMed Central PMCID: PMCPMC3017321. eng.
- 6 14. Padgett PK, Jacobs JV, Kasser SL. Is the BESTest at its best? A suggested brief
7 version based on interrater reliability, validity, internal consistency, and
8 theoretical construct. *Phys Ther.* 2012 Sep;92(9):1197-207. doi:
9 10.2522/ptj.20120056. PubMed PMID: 22677295; eng.
- 10 15. Jácome C, Cruz J, Oliveira A, et al. Validity, Reliability, and Ability to Identify
11 Fall Status of the Berg Balance Scale, BESTest, Mini-BESTest, and Brief-
12 BESTest in Patients With COPD. *Phys Ther.* 2016 Apr 14;96(11):1807-1815. doi:
13 10.2522/ptj.20150391. PubMed PMID: 27081201; Eng.
- 14 16. Franchignoni F, Horak F, Godi M, et al. Using psychometric techniques to
15 improve the Balance Evaluation Systems Test: the mini-BESTest. *J Rehabil Med.*
16 2010 Apr;42(4):323-31. doi: 10.2340/16501977-0537. PubMed PMID:
17 20461334; PubMed Central PMCID: PMCPMC3228839. eng.
- 18 17. Duncan RP, Leddy AL, Cavanaugh JT, et al. Comparative utility of the BESTest,
19 mini-BESTest, and brief-BESTest for predicting falls in individuals with
20 Parkinson disease: a cohort study. *Phys Ther.* 2013 Apr;93(4):542-50. doi:
21 10.2522/ptj.20120302. PubMed PMID: 23174567; PubMed Central PMCID:
22 PMCPMC3613340. eng.
- 23 18. Godi M, Franchignoni F, Caligari M, et al. Comparison of reliability, validity, and
24 responsiveness of the mini-BESTest and Berg Balance Scale in patients with

balance disorders. *Phys Ther.* 2013 Feb;93(2):158-67. doi: 10.2522/ptj.20120171.
PubMed PMID: 23023812; eng.

19. Kottner J, Audigé L, Brorson S, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol.* 2011;64(1):96-106. doi: 10.1016/j.jclinepi.2010.03.002.

20. Terwee CB, Mokkink LB, Knol DL, et al. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res.* 2012 May;21(4):651-7. doi: 10.1007/s11136-011-9960-1. PubMed PMID: 21732199; PubMed Central PMCID: PMCPMC3323819. eng.

21. von Elm E, Altman DG, Egger M, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *BMJ.* 2007 Oct 20;335(7624):806-8. doi: 10.1136/bmj.39335.541782.AD. PubMed PMID: 17947786; PubMed Central PMCID: PMCPMC2034723. eng.

22. Biderman A, Cwikel J, Fried A, et al. Depression and falls among community dwelling elderly people: a search for common risk factors. *J Epidemiol Community Health.* 2002;56(8):631-636. doi: 10.1136/jech.56.8.631. PubMed PMID: PMC1732215.

23. Zhang F, Ferrucci L, Culham E, et al. Performance on five times sit-to-stand task as a predictor of subsequent falls and disability in older persons. *J Aging Health.* 2013 Apr;25(3):478-92. doi: 10.1177/0898264313475813. PubMed PMID: 23407343; eng.

- 1 24. Myers AM, Fletcher PC, Myers AH, et al. Discriminative and evaluative
2 properties of the activities-specific balance confidence (ABC) scale. *J Gerontol A*
3 *Biol Sci Med Sci*. 1998 Jul;53(4):M287-94. PubMed PMID: 18314568; eng.
- 4 25. Branco PS. Validation of the portuguese version of the “Activities-specific
5 Balance Confidence Scale”. *RSPMFR*. 2010;19(2):20-25.
- 6 26. Tsang CS, Liao LR, Chung RC, et al. Psychometric properties of the Mini-
7 Balance Evaluation Systems Test (Mini-BESTest) in community-dwelling
8 individuals with chronic stroke. *Phys Ther*. 2013 Aug;93(8):1102-15. doi:
9 10.2522/ptj.20120454. PubMed PMID: 23559522; eng.
- 10 27. Marques A, Almeida S, Carvalho J, et al. Reliability, Validity, and Ability to
11 Identify Fall Status of the Balance Evaluation Systems Test, Mini-Balance
12 Evaluation Systems Test, and Brief-Balance Evaluation Systems Test in Older
13 People Living in the Community. *Arch Phys Med Rehabil*. 2016;97(12):2166-
14 2173.e1.
- 15 28. Troiano RP, Berrigan D, Dodd KW, et al. Physical activity in the United States
16 measured by accelerometer. *Med Sci Sports Exerc*. 2008 Jan;40(1):181-8. doi:
17 10.1249/mss.0b013e31815a51b3. PubMed PMID: 18091006; eng.
- 18 29. Bonett DG. Sample size requirements for estimating intraclass correlations with
19 desired precision. *Stat Med*. 2002 May 15;21(9):1331-5. doi: 10.1002/sim.1108.
20 PubMed PMID: 12111881; eng.
- 21 30. Greenwood SA, Lindup H, Taylor K, et al. Evaluation of a pragmatic exercise
22 rehabilitation programme in chronic kidney disease. *Nephrol Dial Transplant*.
23 2012 Oct;27 Suppl 3:iii126-34. doi: 10.1093/ndt/gfs272. PubMed PMID:
24 22785111; eng.

- 1 31. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an
2 illustration of appropriate statistical analyses. Clin Rehabil. 1998 Jun;12(3):187-
3 99. PubMed PMID: 9688034; eng.
- 4 32. Fleiss J. Reliability of measurements. In: Fleiss J, editor. The design and analysis
5 of clinical experiments. 1 ed. New York: John Wiley & Sons; 1986.
- 6 33. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an
7 illustration of appropriate statistical analyses. Clin Rehabil. 1998;12(3):187-199.
- 8 34. Weir JP. Quantifying test-retest reliability using the intraclass correlation
9 coefficient and the SEM. J Strength Cond Res. 2005 Feb;19(1):231-40. doi:
10 10.1519/15184.1. PubMed PMID: 15705040; eng.
- 11 35. Huang S-LH, Ching-Lin; Wu, Ruey-Meei; Tai, Chun-Hwei; Lin, Chin-Hsien; Lu,
12 Wen-Shian. Minimal Detectable Change of the Timed "Up & Go" Test and the
13 Dynamic Gait Index in People With Parkinson Disease. Phys Ther.
14 2011;91(1):114-121.
- 15 36. Smidt N, van der Windt DA, Assendelft WJ, et al. Interobserver reproducibility
16 of the assessment of severity of complaints, grip strength, and pressure pain
17 threshold in patients with lateral epicondylitis. Arch Phys Med Rehabil. 2002
18 Aug;83(8):1145-50. PubMed PMID: 12161838; eng.
- 19 37. O'Hoski S, Sibley KM, Brooks D, et al. Construct validity of the BESTest, mini-
20 BESTest and briefBESTest in adults aged 50 years and older. Gait Posture. 2015
21 Sep;42(3):301-5. doi: 10.1016/j.gaitpost.2015.06.006. PubMed PMID:
22 26183191; eng.
- 23 38. Schumacher J, Pientka L, Trampisch U, et al. The prevalence of falls in adults
24 aged 40 years or older in an urban, German population. Results from a telephone

survey. Z Gerontol Geriatr. 2014 Feb;47(2):141-6. doi: 10.1007/s00391-013-0503-y. PubMed PMID: 23743881; eng.

39. Jahn K, Zwergal A, Schniepp R. Gait Disturbances in Old Age: Classification, Diagnosis, and Treatment From a Neurological Perspective. Deutsches Ärzteblatt International. 2010 04/30

11/09/received

01/07/accepted;107(17):306-316. doi: 10.3238/arztebl.2010.0306. PubMed PMID: PMC2872829.

40. Marques A, Silva A, Oliveira A, et al. Validity and Relative Ability of 4 Balance Tests to Identify Fall Status of Older Adults With Type 2 Diabetes. J Geriatr Phys Ther. 2016 Nov 07. doi: 10.1519/jpt.0000000000000109. PubMed PMID: 27824659; eng.

41. Chan AC, Pang MY. Assessing Balance Function in Patients With Total Knee Arthroplasty. Phys Ther. 2015 Apr 16;95(10):1397-407. doi: 10.2522/ptj.20140486. PubMed PMID: 25882482; Eng.

42. Marques A, Cruz J, Quina S, et al. Reliability and Minimal Detectable Change of the Timed Up & Go and the 10-Meter Walk tests in Patients with COPD. COPD. 2016;13(3):279-87.

Table 1 - Participants' characteristics (n=74).

Characteristics	Total (n=74)	Without history of falls (n=46)	With history of falls (n=28)	p-value
Age (years)	63.9 (15.1)	61 (15.8)	68.8 (12.8)	0.029*
Male, n(%)	49 (66.2)	32 (69.6)	17 (60.7)	0.298
BMI (Kg/m ²)	25.1 (4.3)	25.8 (4.7)	24.1 (3.2)	0.068
Walking aid, n(%)	12 (16.2)	5(17.9)	7(15.2)	0.765
Comorbidities, M[IQR]	1 [0, 2]	1 [0, 2]	2 [0, 2.75]	0.403
Activities-specific Balance Confidence scale	64 (26.2)	69.5 (24.8)	55.4 (26.6)	0.034*
BESTest	77.9 (17.3)	82 (15.6)	71.4 (18.1)	0.013*
Mini-BESTest	21.7 (6)	22.8 (5.4)	20 (6.5)	0.061
Brief-BESTest	14.9 (6.4)	16.6 (6.3)	12.1 (5.8)	0.003*

Note: values show mean (SD) unless otherwise indicated. Abbreviations: BESTest, Balance Evaluation Systems Test; BMI, body mass index; IQR, interquartile range; M, median.

Table 2 – Balance tests scores per age ranges (n=74).

Characteristics	30-49 yrs (n=16)	50-69 yrs (n=27)	>70 yrs (n=31)	p-value
BESTest	93.5 (6.6)	76.1 (18.8)	70.8 (14.5)	<0.001
Mini-BESTest	26.2 (1.8)	21.4 (6.4)	19.6 (5.9)	0.001
Brief-BESTest	21.5 (2.8)	14.7 (5.6)	11.5 (5.8)	<0.001

Note: values show mean (SD). Abbreviations: BESTest, Balance Evaluation Systems Test.

1 Table 3 - Test-retest reliability of the Balance Evaluation Systems Test (BESTest), Mini-
2 BESTest and the Brief-BESTest (n=31).

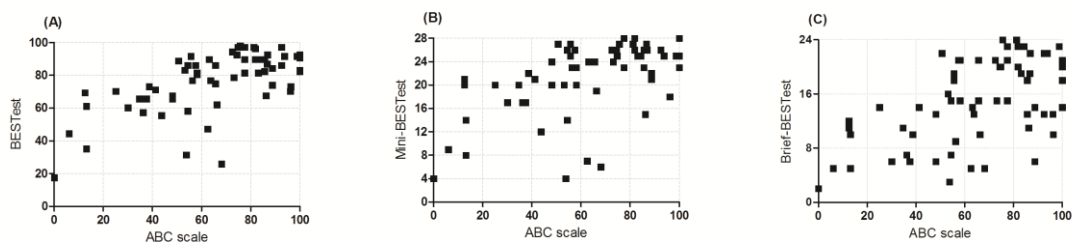
Balance test	Relative reliability		Agreement	
	ICC_{2,1}	95% CI	Mean difference (SD)	95% LA
BESTest	0.94	0.87→0.97	-0.7 (5.9)	-12.3→10.8
Mini-BESTest	0.84	0.70→0.92	-0.4 (2.9)	-6→5.3
Brief-BESTest	0.84	0.69→0.92	0.2 (3)	-5.7→6.1

3 Note: The ICC equations used were the ICC_{2,1}. Abbreviations: 95% CI, 95% confidence
4 intervals; ICC, Intraclass Correlation Coefficient; SD, standard deviation; 95% LA, 95%
5 limits of agreement.

6

1 **Figures**

2 Figure 1 - Scatterplots showing the relationship between the Activities-specific Balance
3 Confidence (ABC) scale and (A) the Balance Evaluation Systems Test (BESTest), (B)
4 the Mini-BESTest and (C) the Brief-BESTest (n=74).



5
6 Figure 2 - Bland and Altman plots of the (A) Balance Evaluation Systems Test (BESTest),
7 (B) Mini-BESTest and (C) Brief-BESTest between two sessions (n=31). The bold line
8 represents the mean difference between sessions 1 and 2 and the dotted lines the 95%
9 limits of agreement.

